Probability and Statistical Applications – Estimation theory

Chapter · November 2014

DOI: 10.13140/2.1.1411.7445

1 author:

Nimmagadda Venkata Nagendram
Acharya Nagarjuna University

Probability and Statistical Applications

By Dr. N V Nagendram

UNIT – I

Probability Theory: Sample spaces Events & Probability; Discrete Probability; Union, intersection and compliments of Events; Conditional Probability; Baye's Theorem .

UNIT – II

Random Variables and Distribution; Random variables Discrete Probability Distributions, continuous probability distribution, Mathematical Expectation or Expectation Binomial, Poisson, Normal, Sampling distribution; Populations and samples, sums and differences. Central limit Elements. Theorem and related applications.

UNIT – III

Estimation – Point estimation, interval estimation, Bayesian estimation, Text of hypothesis, one tail, two tail test, test of Hypothesis concerning means. Test of Hypothesis concerning proportions, F-test, goodness of fit.

UNIT – IV

Linear correlation coefficient Linear regression; Non-linear regression least square fit; Polynomial and curve fittings.

UNIT – V

Queing theory – Markov Chains – Introduction to Queing systems- Elements of a Queuing model – Exponential distribution – Pure birth and death models. Generalized Poisson Queuing model – specialized Poisson Queues.

Text Book: Probability and Statistics By T K V Iyengar S chand, 3rd Edition, 2011.

References:

1. Higher engg. Mathematics by B V Ramana, 2009 Edition.
2. Fundamentals of Mathematical Statistics by S C Gupta & V K Kapoor Sultan Chand & Sons, New Delhi 2009.
3. Probability & Statistics by Schaum outline series, Lipschutz Seymour,TMH,New Delhi 3rd Edition 2009.
4. Probability & Statistics by Miller and freaud, Prentice Hall India, Delhi 7th Edition 2009.

Planned Topics

UNIT – III

Estimation – Point estimation, interval estimation, Bayesian estimation, Tests of hypothesis, one tail, two tail test, test of Hypothesis concerning means. Test of Hypothesis concerning proportions, F-test, goodness of fit.

1. Estimation - Introduction
2. Point Estimation – Definitions, Examples
3. Interval Estimation – Definitions, Examples
4. Problems
5. Bayesian Estimation
6. Problems
7. Tests of Hypothesis – Definitions, Examples
8. Tutorial
9. One tail test, Two tail test – definitions, examples
10. Test of hypothesis concerning means
11. Problems
12. Test of hypothesis concerning proportions
13. Problems
14. F-test
15. Goodness of fit
16. Tutorial

Problem#1 A district official intends to use the mean of a random sample of 150 sixth graders from a very large school district to estimate the mean score which all the sixth graders in the district would get if they took a certain arithmetic achievement test. If based on experience, the official knows that $\sigma = 9.4$ for such data, what can she assert with probability 0.95 about the maximum error? [Ans. 1.50]

Problem #2: If random sample of size n = 20 from a normal population with the variance $\sigma^2 = 225$ has the mean $\bar{x} = 64.3$ construct a 95% confidence interval for the population mean μ?

[Ans. $57.7 < \mu < 70.9$]

Problem #3: An industrial designer wants to determine the average amount of time it takes an adult to assemble an "easy to assemble" toy. Use the following data (in minutes) as random sample to construct a 95% confidence interval for the mean of the population sample:

17	13	18	19	17	21	29	22	16	28	21	15
26	23	24	20	8	17	17	21	32	18	25	22
16	10	20	22	19	14	30	22	12	24	28	11

[Ans. $18.05 < \mu < 21.79$]

Problem #4: A paint manufacturer wants to determine the average drying time of a new interior wall paint. If for 12 test areas of equal size he obtained a mean drying time of 66.3 minutes and a standard deviation of 8.4 minutes, construct a 95% confidence interval for the true mean μ? [Ans. $61.0 < \mu < 71.6$]

Problem #5: The idea of a college wants to use the mean of a random sample to estimate the average amount of the students take to get from one class to the next, and she wants to be able to assert with 99% confidence that the error is at most 0.25 minute? [Ans. 208]

Problem #6: suppose that it is known from experience that the standard deviation of the weight of 8-ounce packages of cookies made by a certain bakery is 0.16 ounce. To check whether its production is under control on a given day, namely, to check whether the true average weight of the packages is 8 ounces, employees select a random sample of 25 packages and find that their mean weight is $\bar{x} = 8.091$ ounces. Since the bakery stands to lose money when $\mu > 8$ and the customer loses out when $\mu < 8$, test the null hypothesis $\mu = 8$ against the attentive hypothesis is $\mu \neq 8$ at the 0.01 level of significance? [Ans. $2.84 > 2.575$]

Problem#1 A district official intends to use the mean of a random sample of 150 sixth graders from a very large school district to estimate the mean score which all the sixth graders in the district would get if they took a certain arithmetic achievement test. If based on experience, the official knows that $\sigma = 9.4$ for such data, what can she assert with probability 0.95 about the maximum error? [Ans. 1.50]

Solution: $n = 150$, $\sigma = 9.4$, $\alpha = 0.05$, $Z_{0.025} = 1.96$

$$\text{Max Error} = 1.96 \text{ X } \frac{9.4}{\sqrt{150}} \equiv 1.50$$

Hence the solution.

Problem #2: If random sample of size n = 20 from a normal population with the variance $\sigma^2 = 225$ has the mean $\bar{x} = 64.3$ construct a 95% confidence interval for the population mean μ?

[Ans. $57.7 < \mu < 70.9$]

Solution: Substituting $n = 20$, $\sigma = 15$, $\bar{x} \equiv 64.3$, $\alpha = 0.05$, $Z_{0.025} = 1.96$ into the confidence interval formula of μ, we get $64.3 - 1.96\, X \frac{15}{\sqrt{20}} < \mu < 64.3 \pm 1.96\, X \frac{15}{\sqrt{20}}$ which reduces to

$57.7 < \mu < 70.9$. Hence the solution.

Problem #3: An industrial designer wants to determine the average amount of time it takes an adult to assemble an "easy to assemble" toy. Use the following data (in minutes) as random sample to construct a 95% confidence interval for the mean of the population sample:

17	13	18	19	17	21	29	22	16	28	21	15
26	23	24	20	8	17	17	21	32	18	25	22
16	10	20	22	19	14	30	22	12	24	28	11

[Ans. $18.05 < \mu < 21.79$]

Solution: Substituting $n = 36$, $s = 5.73$, $\bar{x} \equiv 19.92$, $\alpha = 0.05$, $Z_{0.025} = 1.96$ for σ into the confidence interval formula form, we get $19.92 - 1.96\, X \frac{5.73}{\sqrt{36}} < \mu < 19.92 \pm 1.96\, X \frac{5.73}{\sqrt{36}}$.

Thus, the 95% confidence limits are 18.05, 21.79 minutes. Hence the solution.

Problem #4: A paint manufacturer wants to determine the average drying time of a new interior wall paint. If for 12 test areas of equal size he obtained a mean drying time of 66.3 minutes and a standard deviation of 8.4 minutes, construct a 95% confidence interval for the true mean μ? [Ans. $61.0 < \mu < 71.6$]

Solution: Substituting n = 12, s = 8.4, $\bar{x} \equiv 66.3$, α = 0.05, $t_{0.025,11} = 2.201$, the 95% confidence interval μ becomes, we get $66..3 \ - 2.201 X \dfrac{8.4}{\sqrt{12}} < \mu < 66.3 \pm 2.201 X \dfrac{8.4}{\sqrt{12}}$.

Thus, the 95% confidence limits are 61.0, 71.6. this means that we can assert with 95% confidence that interval from 61.0 minutes to 71.6 minutes contains the true average drying time of the paint. Hence the solution.

Problem #5: The dean of a college wants to use the mean of a random sample to estimate the average amount of the students take to get from one class to the next, and she wants to be able to assert with 99% confidence that the error is at most 0.25 minute? it can be presumed from experience that σ = 1.40 how large a sample will she have to take? [Ans. 208]

Solution: σ = 1.40 and $Z_{0.005} = 2.575$, E = 0.25 into the formula for n we get

$$n=\left(\frac{2.575 X 1.40}{0.25}\right)^2 \approx 207.9 \approx 208.$$ Hence the solution.

Problem #6: suppose that it is known from experience that the standard deviation of the weight of 8-ounce packages of cookies made by a certain bakery is 0.16 ounce. To check whether its production is under control on a given day, namely, to check whether the true average weight of the packages is 8 ounces, employees select a random sample of 25 packages and find that their mean weight is $\bar{x}$ = 8.091 ounces. Since the bakery stands to lose money when μ > 8 and the customer loses out when μ < 8, test the null hypothesis μ = 8 against the attentive hypothesis is μ ≠ 8 at the 0.01 level of significance? [Ans. 2.84 > 2.575]

Solution: 1. $H_0 : \mu = 8$; $H_1 : \mu \neq 8$

2. α = 0.01

3. reject the null hypothesis if Z ≤ -2.575 or Z ≥ 2.575

4. $\bar{x}$ = 8.091, $\mu_0 = 8$, σ = 0.16 and n = 25

$$Z = \frac{8.091-8}{0.16/\sqrt{25}} \approx 2.84.$$ Hence the solution.

Problem #1 In six determinations of the melting point of tin, a chemist obtained a mean of 232.26 degrees Celsius with a standard deviation σ of 0.14 degree. If he uses this mean as the actual melting point of tin, what can the chemist assert with 98% confidence about the maximum error? [Ans. E= 0.19 degree]

Solution: given n = 6, s = 0.14, $t_{\alpha/2} = t_{0.01} = 3.365$

we have, E = $t_{\alpha/2} \frac{s}{\sqrt{n}} = \frac{3.365 \ X \ 0.14}{\sqrt{6}} \equiv 0.19$ Hence the solution.

Problem #2 A research worker wants to determine the average that it takes a mechanic to rotate the tyres of a car, and he want to be able to assert with 95% confidence that the mean of his sample is off by at most 0.50 minute. If he can presume from past experience that σ = 1.6 minutes, how large a sample will he has to take? [Ans. 39.3 ≈ 40]

Solution: Given σ = 1.6, E = 0.50, $Z_{\alpha/2} = Z_{0.025} = 1.96$

we have, n = $\left(Z_{\alpha/2} \frac{\sigma}{E}\right)^2 = \left(\frac{1.96 \ X 1.6}{0.50}\right)^2 \equiv 39.3 \approx 40$ Hence the solution.

Problem #3 An industrial engineer intends to use the mean of a random sample of size n = 150 to estimate the average mechanical aptitude as measured by a certain test of assembly line workers in a large industry. If on the basis of experience , the engineer can assume that σ = 6.2 for such data, who can he assert with probability 0.99 about the maximum size of his error?

[Ans: P(0.99) error 1.30]

Solution: Given n = 150, σ = 6.2, $Z_{\alpha/2} = Z_{0.005} = 2.575$

we have, E = $\left(Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = \left(\frac{2.575 \ X \ 6.2}{\sqrt{150}}\right) \equiv 1.30$ Hence the solution.

Problem #4 Measurement of the blood pressure of 25 elderly women have a mean of $\bar{x}$ =140 mm of mercury. If these data can be looked upon as a random sample from a normal population with σ = 10 mm of mercury, construct a 95% confidence interval for the population mean μ ? [Ans. 136.08 < μ < 143.92]

Solution: Given $\bar{x}$ =140 mm, n = 25, σ = 10, $Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$

The 95% confidence interval for the population mean μ is,

$$\bar{x} - 1.96 \ X \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} \pm 1.96 \ X \frac{\sigma}{\sqrt{n}} = 140 - 1.96 \ X \frac{10}{\sqrt{25}} < \mu < 140 \pm 1.96 \ X \frac{10}{\sqrt{25}}$$

$$= 140 - 1.96 \ X \ 2 < \mu < 140 \pm 1.96 \ X \ 2 = 140 - 3.92 < \mu < 140 \pm 3.92$$

$= 136.08 - 3.92 < \mu < 143.92$. Hence the solution.

Problem #5 A random sample of size n = 100 is taken from a population with σ = 5.1. given that the sample mean is $\bar{x}$ = 21.6, construct a 95% confidence interval for the population mean μ ? [Ans. 20.6 < μ < 22.6]

Solution: Given $\bar{x}$ =21.6, n = 100, σ = 5.1, $Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$

The 95% confidence interval for the population mean μ is,

$$\bar{x} - 1.96\ X \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} \pm 1.96\ X \frac{\sigma}{\sqrt{n}} = 21.6 - 1.96\ X \frac{5.1}{\sqrt{100}} < \mu < 140 \pm 21.6\ X \frac{5.1}{\sqrt{100}}$$

$$= 21.6 - 1.96\ X\ 0.51 < \mu < 21.6 \pm 1.96\ X\ 0.51 = 20.6 < \mu < 22.6$$

It is 95% confident that the interval (20.6, 22.6) contains the population mean μ.

Hence the solution.

Problem #6: If a random sample of size n = 20 from a normal population with the variance σ^2 = 225 has the mean x = 64.3 , construct a 95% confidence interval for the mean population mean μ? [Ans. 57.7 < μ < 70.9]

Solution: Given $\bar{x}$ =64.3, n = 20, σ = ± 15, α = 0.05, $Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$

The 95% confidence interval for the population mean μ is,

$$\bar{x} - 1.96\ X \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} \pm 1.96\ X \frac{\sigma}{\sqrt{n}} = 64.3 - 1.96\ X \frac{15}{\sqrt{20}} < \mu < 64.3 \pm 1.96\ X \frac{15}{\sqrt{20}}$$

$$= 64.3 - 6.60 < \mu < 64.3 \pm 6.6 = 57.7 < \mu < 70.9$$

It is 95% confident that the interval (57.7, 70.9) contains the population mean μ.

Hence the solution.

Problem #1 A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance? [Ans. |Z| = 19.17]

Problem #2 a die tossed 960 times and it falls with 5 upwards 184 times. Is the die unbiased at a level of significance of 0.01? [Ans. |Z| < 2.58]

Problem #3 A die tossed 256 times and it turns up with an even digit 150 times. Is the die biased? [Ans. |Z| > 1.96]

Problem #4: A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance? [Ans: |Z| = 1.6 < 1.96]

Problem #1 A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance? [Ans. |Z| = 19.17]

Solution: here n = 960, p = probability (H) = $\frac{1}{2}$

q = 1 – p = $\frac{1}{2}$ and μ = np = 960 X $\frac{1}{2}$ = 480

$\sigma = \sqrt{npq} = \sqrt{(np)q} = \sqrt{480X1/2} = \sqrt{240} = 15.49$

X = number of successes = 183

1. Null hypothesis H_0 : the coin is unbiased
2. Alternate hypothesis H_1: the coin is biased
3. level of significance , α = 0.05
4. the test statistic = Z = $\frac{\bar{x}-\mu}{\sigma} \equiv \frac{183-480}{15.49} \equiv \frac{-297}{15.49} \equiv -19.17$

So |Z| = 19.17 as |Z| > 1.96, the null hypothesis H_0 has to be rejected at 5% level of significance and we conclude that the coin is biased. Hence the solution.

Problem #2 a die tossed 960 times and it falls with 5 upwards 184 times. Is the die unbiased at a level of significance of 0.01? [Ans. |Z| < 2.58]

Solution: here n = 960, p = probability of throwing 5 with one die = $\frac{1}{6}$

q = 1 – p = $\frac{5}{6}$ and μ = np = 960 X $\frac{1}{6}$ = 160

$\sigma = \sqrt{npq} = \sqrt{(np)q} = \sqrt{160X5/6} =$ 11.55

$\bar{x}$ = number of successes = 184

1. Null hypothesis H_0 : the coin is unbiased
2. Alternate hypothesis H_1: the coin is biased
3. level of significance , α = 0.01
4. the test statistic = Z = $\frac{\bar{x}-\mu}{\sigma} \equiv \frac{184-160}{11.55} \equiv \frac{24}{11.55} \equiv 2.08$

So |Z| < 2.08 the null hypothesis H_0 has to be accepted at 1% level of significance and we conclude that the die is unbiased. Hence the solution.

Problem #3 A die tossed 256 times and it turns up with an even digit 150 times. Is the die biased? [Ans. |Z| > 1.96]

Solution: here n = 256, p = probability of getting an even digit (2 or 4 or 6) = $\frac{3}{6} \equiv \frac{1}{2}$

q = 1 – p = $\frac{1}{2}$ and μ = np = 256 X $\frac{1}{2}$ = 128

σ = $\sqrt{npq}$ = $\sqrt{(np)q}$ = $\sqrt{128X1/2} \equiv \sqrt{64} \equiv 8$

$\bar{x}$ = number of successes = 150

1. Null hypothesis H_0 : the die is unbiased
2. Alternate hypothesis H_1: the die is biased
3. level of significance , α = 0.05
4. the test statistic = Z = $\frac{\bar{x}-\mu}{\sigma} \equiv \frac{150-128}{8} \equiv \frac{22}{8} \equiv 2.75$

So |Z| > 1.96, the null hypothesis H_0 has to be rejected at 5% level of significance and we conclude that the die is unbiased. Hence the solution.

Problem #4: A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance? [Ans: |Z| = 1.6 < 1.96]

Solution: here n = 400, p = probability of getting head = $\frac{1}{2}$

q = 1 – p = $\frac{1}{2}$ and μ = np = 400 X $\frac{1}{2}$ = 200

σ = $\sqrt{npq}$ = $\sqrt{(np)q}$ = $\sqrt{200X1/2} \equiv \sqrt{100} \equiv 10$

$\bar{x}$ = number of successes = 216

1. Null hypothesis H_0 : the die is unbiased
2. Alternate hypothesis H_1: the die is biased
3. level of significance , α = 0.05
4. the test statistic = Z = $\frac{\bar{x}-\mu}{\sigma} \equiv \frac{216-200}{10} \equiv \frac{16}{10} \equiv 1.6$

So |Z| < 1.96, the null hypothesis H_0 has to be accepted and we conclude that the coin is unbiased. Hence the solution.

Chapter 3 Estimation Theory Tutorial 4

Problems Vs Solutions **By Dr. N V Nagendram**

Problem #1 The following sample data pertain to the shipments received by a large firm from three different vendors.

Vendor	Number rejected	Number imperfect But acceptable	Number perfect
A	12	23	89
B	8	12	62
C	21	30	119

Test at 0.01 level of significance whether the three vendors ship products of equality?

[Ans. 1.27<13.277]

Problem #2 In a random sample, the weights of 24 brown angus steers of a certain age have a s.d of 238 pounds. Assuming that the weights constitute a random sample from a normal population, test $\sigma = 250$ pounds against the two sided alternative $\sigma \neq 250$ pounds at the 0.01 level of significance? [Ans. 9.26 < 20.84 < 44.181]

Problem #3 In a random sample, the weights of 24 brown angus steers of a certain age have a s.d of 238 pounds. Assuming that the weights constitute a random sample from a normal population, test $\sigma = 250$ pounds against the two sided alternative $\sigma \neq 250$ pounds at the 0.02 level of significance whether it is reasonable to assume that the two populations sampled have equal variances ? [Ans. 1.8 < 11]

Problem #4 Nine determinations of the specific heat of iron had a s.d of 0.0086. Assuming that the determinations constitute a random sample from a normal population, test $H_0 : \sigma = 0.01$ against $H_1 : \sigma < 0.01$ at 0.05 level of significance ? [Ans. 5.92 > 2.733]

Problem #5 The following are the average weekly losses of work hours due to accidents in 10 industrial plants before and after a certain safety program was put into operation 45 and 36, 73 and 60, 46 and 44, 124 and 119,33 and 35,57 and 51,83 and 77,34 and 29,26 and 24, 17 and 11. Test whether the safety program is effective or not?[Ans. 4.03> 1.833 is effective]

Problem #6 To compare two kinds of bumper gaurds, six of each kind were mounted on a certain make of compact car. Then each car was run into a wall at S miles per Hour, and the following are the costs of the repairs (into)

Bumper Guard - 1	127	168	143	165	122	139
Bumper Guard - 2	154	135	132	171	153	149

Test whether the difference between the means of these two samples significant?

[Ans. -3.169< -0.52 < 3.19]

Problem #7 To find out whether the inhabitants of two islands may be regarded as having the same racial ancestry, an anthropologist determines the cephalic indices of six adult males from each island, getting $\overline{x_1}$ = 77.4 and $\overline{x_2}$ = 72.2 the corresponding standard deviations σ_1 = 3.3 and σ_2 = 2.1. Test whether the difference between the two samples means can reasonably be attributed to chance? Assume the populations sampled the two samples and have equal variances. [Ans. 3.26 > 3.169]

Problem #8 A study of the number launches that executives in the insurance and banking industries claim as deductible expenses per month was based on random samples and yielded the following results.

N_1= 40	$\overline{x_1}$ = 9.1	s_1 = 1.9
N_2 = 50	$\overline{x_2}$ = 8.0	s_2 = 2.1

Test the null hypothesis μ_1 - μ_2 against the alternative hypothesis $\mu_1 - \mu_2 \neq 0$? **[Ans. 3.26 > 3.169]**

Problem #9 Five measurements of the tar content of a certain kind of cigarette yielded 14.5, 14.2, 14.3 and 14.6 mg/cigarette. Assuming that the data are a random sample from a normal population, show that at the 0,05 level of significance the null hypothesis μ = 14 must be rejected in favour of the alternative $\mu \neq 14$? [Ans. 5.66 > 2.776]

Problem #10 According to the rules established for a reading comprehension test, 5 th graders should average 84.3 with a standard deviation of 8.6, if 45 randomly selected 5th graders from a certain school averaged 87.8, use the four steps to test the null hypothesis μ = 84.3 against the alternative μ > 84.3 at the 0.01 level of significance ? [Ans. 2.73 > 2.33]

Problem #11 Determine the goodness of fit of the data given below:

Height (inches)	Number of students
60 - 62	5
63 – 65	18
66 - 68	42
69 – 71	27
72 - 74	8
Total →	100

Here consider mean μ = 67.45, s.d = s = 2.92 inches. [Ans. 0.959 > 0.103]

Problem #12 The table below shows the observed frequencies in tossing a die 120 times can we consider the die fair?

Faces	1	2	3	4	5	6
Observed frequency O_i	25	17	15	23	24	16

[Ans.$\chi^2 < \chi^2_{0.05,\,5}$]

Problem #13 Test whether there is significant difference at 0.05 level of significance in the quality of manufacturing process of four companies A,B,C,D if the number of defectives are 26, 23, 15, 32 respectively Assume that each company has a productions of 200 items?

[Ans. .$\chi^2 = 7.10 < \chi^2_{0.05,\,3} = 7.815$]

Problem #14 To determine the effectiveness of drugs against a disease, three types of drugs (from three different drug manufacturing companies) were tested on 50 persons with the following results

		Drug A	Drug B	Drug C	Total
	No relief	11	13	9	33
Effectiveness	Some relief	32	28	27	87
	Total relief	7	9	14	30
	Total →	50	50	50	150

[Ans. .$\chi^2 = 3.81 < \chi^2_{0.05} = 9.488$]

Problem #15 In order to determine whether "perfection" in job depends on the "experience ", 400 persons were examined yielding the following data

		High experience	medium experience	low experience	Total experience
	Excellent	23	60	29	112
Effectiveness	Good	28	79	60	167
	Satisfactory	9	49	63	121
	Total →	60	188	152	400

[Ans. .$\chi^2 = 20.34 > \chi^2_{0.01} = 13.277$]

Problem #16 Test for goodness of fit of a bionomial distribution to the data given below:

X_i	0	1	2	3	4	5	6
O_i	5	18	28	12	7	6	4

[Ans. .$\chi^2 = 6.39 > \chi^2_{0.05,2} = 5.99$]

Problem #16 A set of 5 identical coins is tossed 320 times and the number of heads appearing each time is recorded. The results are: test 5% level of significance whether B.Population?

No. of heads	0	1	2	3	4	5
Frequency	5	18	28	12	7	6

[Ans. .$\chi^2 = 9.053 < \chi^2_{0.05,4} = 9.488$]

Problem #17 The following data shows suicides of women in 8 ferman states during 14 years.

No. of Suicides in a state/Yr.	0	1	2	3	4	5	6	7	Total
Observed Freq.	364	376	218	89	33	13	2	1	1096

Test at the 5% level of significance whether the data is from a Poisson population?

[Ans. .$\chi^2 = 9.99 < \chi^2_{0.05,4} = 9.49$]

Problem #18 use the dat in the following table to test at the 0.01 level of significance whether a person's ability in mathematics is independent of his or her interest in statistics.

		Ability in Mathematics		
		Low	Average	High
Interest	Low	63	42	15
In	Average	18	61	31
Statistics	High	14	47	29

[Ans. .$\chi^2 = 32.14 > \chi^2_{0.01} = 13.27$]

Problem #20 samples of three kinds of materials, subjected to extreme temperature changes, produced this results shown in the following table:

	Material			Total
	A	B	C	
Crumbled	41	27	22	90
Remained intact	79	53	78	210
Total →				300

[Ans. .$\chi^2 = 4.75$ does not exceed $\chi^2_{0.05} = 5.91$]

Problem #21 An oil company claims that less than 20% of all car owners have not tried its gasoline. Test this claim at the 0.01 level of significance, if a random check reveals that 22 of 200 car owners have not tried the oil company's gasoline? [Ans. -3.18 < -2.33]

Problem #21 a sample survey at a market showed that 204 of 300 shoppers regularly use cents – off coupons. Use large sample confidence interval to construct a 95% confidence interval for the corresponding true proportion? [Ans. $0.628 < p < 0.732$]

Problem #22 In the comparison of two kinds of paint, a consumer testing service finds that four 1-gallon cans of one brand cover on the average546 square feet with s.d of 31 square feet whereas four 1 gallon cans of another brand coveron the average 492 square feet with s.d of 26 square feet. Assuming that the two populations sampled are normal and have variances. Test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternate hypothesis $\mu_1 - \mu_2 > 0$ at the 0.05 level of significance? [Ans. 2.67 > 1.943]

<<End of tutorial 4 of Estimation Theory Problems >>

Problem #1 Suppose that we want to test on the basis of n = 35 determinants and at the 0.05 level of significance whether the thermal conductivity of a certain kind of cement brick is 0.340, as has been claimed. From information gathered in similar studies, we can expect that the variability of such determinations is given by $\sigma = 0.010$ and suppose that the mean of 35 determinations is 0.343?

Problem #2 A trucking suspects the claim that the average life time of certain tires is atleast 28,000 miles. To check the claim, the firm puts 40 of these tires on its trucks and gets a mean life time of 27,463 miles with a standard deviation of 1,348 miles. What can it conclude if the probability of a type I error is to be at most 0.01?

Problem #3 An article in a journal describes the results of tensile adhesion tests on some alloy specimens. The load at specimen failure is as follows in mPa:

19.8	18.5	17.6	16.7	15.8	15.4	14.1	13.6	11.9	11.4
11.4	8.8	7.5	15.4	15.4	19.5	14.9	12.7	11.9	11.4
10.1	7.9								

The sample mean is $\bar{x} \equiv 13.71$ and the sample standard deviation is s = 3.55 verify both your self. Do the data suggest that the mean load at failure exceeds 10 mPa. Assume that load at failure has a normal distribution and use $\alpha = 0.05$?

Problem #4 A company claims that its light bulbs are superior to those of its main competitor. If a study showed that a sample of $n_1 = 40$ of its bulbs had a mean life time of 647 hours of continuous use with a standard deviation of 27 hours, while a sample of $n_2 = 40$ bulbs made by its main competitor had a mean life time of 638 hours of continuous use with a standard deviation of 31 hours, does this substantiate the claim at the 0.05 level of significance?

Problem #5 A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry. And formulation 2 has a new drying ingredient that should reduce the drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2. The two sample average drying times are $\bar{x}_1 \equiv 121 mm\, and\ \ \bar{x}_2 \equiv 112 mm\, resply.$ what conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

Problem #6 In the comparison of two kinds of paint, a consumer testing service finds that four 1-gallon cans of one brand cover on the average 546 square feet with a standard deviation of 31 square feet, where as four 1-gallon cans of another brand cover on the average 492 dquare feet with a standard deviation of 26 square feet. Assuming that the two populations sampled are normal and have equal variances, test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 > 0$ at the 0.05 level of significance? [Ans. 2.67 > 1.943]

Problem #7 A manufacturer of video display units is testing two micro circuit designs to determine whether they produce equivalent current flow. Development engineering has obtained the following data

Design	n	$\overline{x_i}$	S^2
1	n_1= 15	$\overline{x_1}$ =24.2	$s1^2$
2	n_2= 15	$\overline{x_2}$ = 23.9	$s2^2$

Using α=0.10, determine whether there is any difference in mean current flow between two designs, where both populations are assumed to be normal and with unknown, unequal variances σ_1^2 and σ_2^2 ? [Ans. 0.18 < 1.753]

Problem #8 The following are the average weekly losses of worker hours due to accidents in 10 industrial plants before and after a certain safety program was put into operation.

Safety before	45	73	60	44	119	35	51	77	29	24
After	36	60	46	124	33	57	83	34	26	17

Use 0.05 level of significance to test whether the safety program is effective?

[Ans. 4.03 > 1.833]

Problem #9 In random sample, 136 of 400 persons given a flu vaccine experienced some discomfort. Construct a 95% confidence interval for the true proportion of persons who will experience some discomfort from the vaccine? [Ans. 0.29 < p < 0.39]

[note : Max. Error of estimate E=$Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$]

Problem #10 In a sample survey conducted in a large city 136 of 400 persons answered yes to the question whether their city's public transportation is adequate. With 99% confidence what can be say about the maximum error if $\frac{x}{n} \equiv \frac{136}{400} \equiv 0.34$ is used as an estimate of the corresponding true proportion? [Ans. 0.061]

Problem #1 Suppose that we want to test on the basis of n = 35 determinants and at the 0.05 level of significance whether the thermal conductivity of a certain kind of cement brick is 0.340, as has been claimed. From information gathered in similar studies, we can expect that the variability of such determinations is given by σ = 0.010 and suppose that the mean of 35 determinations is 0.343?

Solution: Here the parameter of interest is the mean of thermal conductivity (μ)

1. $H_0 : \mu = 0.340$ and $H_1 : \mu \neq 0.340$
2. $\alpha = 0.05$
3. Criterion: reject null hypothesis if Z < - 1.96 or Z > 1.96 where $Z = \boxed{\dfrac{\overline{X} - \mu_0}{\left(\sigma/\sqrt{n}\right)}}$
4. **calculations:** since n = 35, σ=0.010, $\overline{X} = 0.343$, $\mu_0 \equiv 0.340$

$$Z = \frac{\overline{X} - \mu_0}{\left(\sigma/\sqrt{n}\right)} = \frac{0.343 - 0.340}{\left(0.010/\sqrt{35}\right)} \equiv 1.77$$

5. decision: since Z = 1.77 < 1.96 hence H_0 CAN'T BE REJECTED.

To find P-value: Z = 1.77 and H_1 is two tailed so P-value is

$$P = 2(1 - F(1.77)) = 2(1 - 0.9616) = 0.07868$$

Therefore $P \geq \alpha = 0.05$ H_0 can't e rejected, agreeing our earlier result.

Hence the solution.

Problem #2 A trucking suspects the claim that the average life time of certain tires is atleast 28,000 miles. To check the claim, the firm puts 40 of these tires on its trucks and gets a mean life time of 27,463 miles with a standard deviation of 1,348 miles. What can it conclude if the probability of a type I error is to be at most 0.01?

Solution: Here the parameter of interest is the life time of certain tires (μ)

1. $H_0 : \mu > 28000$ and $H_1 : \mu < 28000$
2. $\alpha \leq 0.01$
3. Criterion: reject null hypothesis if Z < - 2.33 or Z > 2.33 where $Z = \dfrac{\overline{X} - \mu_0}{\left(\sigma/\sqrt{n}\right)}$ with σ replaced by s since it is of type I error, level of significance is 0.01
4. **calculations:** since n = 40, s=1348, $\overline{X} = 27643$, $\mu_0 \equiv 28000$

$$Z = \frac{\overline{X} - \mu_0}{\left(s/\sqrt{n}\right)} = \frac{27643 - 28000}{\left(1348/\sqrt{40}\right)} \equiv -2.52$$

5. decision: since Z = -2.52 < - 2.33 hence H_0 must be REJECTED at 0.01 level of significance.

Hence the solution.

Problem #3 An article in a journal describes the results of tensile adhesion tests on some alloy specimens. The load at specimen failure is as follows in mPa:

19.8	18.5	17.6	16.7	15.8	15.4	14.1	13.6	11.9	11.4
11.4	8.8	7.5	15.4	15.4	19.5	14.9	12.7	11.9	11.4
10.1	7.9								

The sample mean is $\bar{x} \equiv 13.71$ and the sample standard deviation is s = 3.55 verify both your self. Do the data suggest that the mean load at failure exceeds 10 mPa. Assume that load at failure has a normal distribution and use $\alpha = 0.05$?

Solution: Here the parameter of interest is the mean load at failure (μ)

1. $H_0 : \mu = 10$ and $H_1 : \mu > 10$
2. $\alpha = 0.05$
3. Criterion: reject null hypothesis if $t > t_\alpha = t_{0.05} = 1.721$ with $\upsilon = 22 - 1 = 21$ degrees of freedom where t = $\dfrac{\bar{X} - \mu_0}{\left(\sigma/\sqrt{n}\right)}$
4. **calculations:** since n = 22, s = 3.5, $\bar{X}$ = 13.71, $\mu_0 \equiv 10$

$$t = \frac{\bar{X} - \mu_0}{\left(s/\sqrt{n}\right)} = \frac{13.71 - 10}{\left(3.5/\sqrt{22}\right)} = \frac{3.71}{8.7571} = 4.90$$

5. Decision: since t = 4.90 > $t_{0.05}$ = 1.721. Hence H_0 must be REJECTED at 0.05 level of significance and the mean load failure exceeds 10 mPa.

Hence the solution.

Problem #4 A company claims that its light bulbs are superior to those of its main competitor. If a study showed that a sample of $n_1 = 40$ of its bulbs had a mean life time of 647 hours of continuous use with a standard deviation of 27 hours, while a sample of $n_2 = 40$ bulbs made by its main competitor had a mean life time of 638 hours of continuous use with a standard deviation of 31 hours, does this substantiate the claim at the 0.05 level of significance?

Solution: Here the parameter of interest is the difference in mean life time $\mu_1 - \mu_2$

1. $H_0 : \mu_1 - \mu_2 = 0$ and $H_1 : \mu_1 - \mu_2 > 0$; $\alpha = 0.05$
2. Criterion: reject null hypothesis if $Z > z_\alpha = z_{0.05} = 1.645$ where in Z-statistic use s_1^2 and s_2^2 in-place of σ_1^2 and σ_2^2
3. **calculations:** since $n_1 = 40$, $n_2 = 40$, $\bar{X}_1 = 647$, $\bar{X}_2 = 638$, $s_1 = 27$, $s_2 = 31$, $\delta = 0$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} \pm \dfrac{s_2^2}{n_2}}} = \frac{647 - 638}{\sqrt{\dfrac{26^2}{40} \pm \dfrac{31^2}{40}}} = \frac{9}{6.5217} = 1.38$$

4. Decision: since Z = 1.38 < $z_{0.05}$ = 1.645. Hence H_0 can not be REJECTED at 0.05 level of significance and there is no difference between the two sample means.

Hence the solution.

Problem #5 A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry. And formulation 2 has a new drying ingredient that should reduce the drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2. The two sample average drying times are $\overline{x_1} \equiv 121 mm and \ \overline{x_2} \equiv 112 mm resply.$ what conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

Solution: Here the parameter is Quantity of interest is the difference in mean drying time $\mu_1 - \mu_2$

1. $H_0 : \mu_1 - \mu_2 = 0$ and $H_1 : \mu_1 - \mu_2 > 0$; $\alpha = 0.05$
2. Criterion: Reject null hypothesis if $Z > z_\alpha = z_{0.05} = 1.645$ where in Z-statistic variances are σ_1^2 and σ_2^2
3. **calculations:** since $n_1 = 10$, $n_2 = 10$, $\overline{X}_1 = 121$, $\overline{X}_2 = 112$, $\sigma_1 = 8$, $\sigma_2 = 8$, $\delta = 0$
$$Z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} \pm \frac{\sigma_2^2}{n_2}}} = \frac{121 - 112}{\sqrt{\frac{8^2}{10} \pm \frac{8^2}{10}}} = \frac{9}{3.5714} = 2.52$$
4. Decision: since $Z = 2.52 > z_{0.05} = 1.645$. Hence H_0 must be REJECTED at 0.05 level of significance and that adding noew ingredient to the paint significantly reduces drying time.

Hence the solution.

Problem #6 In the comparison of two kinds of paint, a consumer testing service finds that four 1-gallon cans of one brand cover on the average 546 square feet with a standard deviation of 31 square feet, where as four 1-gallon cans of another brand cover on the average 492 dquare feet with a standard deviation of 26 square feet. Assuming that the two populations sampled are normal and have equal variances, test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 > 0$ at the 0.05 level of significance? [Ans. 2.67 > 1.943]

Solution: Here the parameter is Quantity of interest is the difference in average coverage area by paint $\mu_1 - \mu_2$

1. $H_0 : \mu_1 - \mu_2 = 0$ and $H_1 : \mu_1 - \mu_2 > 0$; $\alpha = 0.05$
2. Criterion: Reject null hypothesis if $t > t_\alpha = t_{0.05} = 1.943$ where the value $t_{0.05} = n_1 + n_2 - 2 = 4 + 4 - 2 = 6$ degree of freedom.
3. **calculations:** since $n_1 = 4$, $n_2 = 4$, $\overline{X}_1 = 546$, $\overline{X}_2 = 492$, $s_1 = 31$, $s_2 = 26$, $\delta = 0$

   we calculate first s_p, as $s_p = \sqrt{\frac{3(31)^2 \pm 3(26)^2}{4 \pm 4 - 2}} = 28.609$

   and then $t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} \pm \frac{\sigma_2^2}{n_2}}} = \frac{546 - 492}{\sqrt{\frac{28.609^2}{4} \pm \frac{28.609^2}{4}}} = \frac{54}{20.2247} = 2.67$

4. **Decision:** since t = 2.67 > $t_{0.05}$ = 1.943. Hence H_0 must be REJECTED at 0.05 level of significance and that the average the first kind of paint covers a greater area than the second. Hence the solution.

Problem #7 A manufacturer of video display units is testing two micro circuit designs to determine whether they produce equivalent current flow. Development engineering has obtained the following data

Design	N	$\overline{x_i}$	S^2
1	n_1= 15	$\overline{x_1}$ =24.2	$s1^2$
2	n_2= 15	$\overline{x_2}$ = 23.9	$s2^2$

Using α=0.10, determine whether there is any difference in mean current flow between two designs, where both populations are assumed to be normal and with unknown, unequal variances σ_1^2 and σ_2^2 ? [Ans. 0.18 < 1.753]

Solution: Here the parameters of interest are the mean current flows for the two circuit designs say μ_1 and μ_2

1. $H_0 : \mu_1 - \mu_2 = 0$ and $H_1 : \mu_1 \neq \mu_2$; $\alpha = 0.10$
2. Criterion: Reject null hypothesis if $t > t_\alpha = t_{0.10}$ = 1.753 or t < - 1.753 is the value of $t_{\alpha/2} = t_{0.05}$ with degrees of freedom υ given by

$$\upsilon = \frac{\left(\frac{10}{15}+\frac{20}{10}\right)^2}{\frac{\left(10/15\right)^2}{14}+\frac{\left(20/10\right)^2}{9}} \equiv 14.9 \approx 15$$

3. **calculations:** since n_1 = 15, n_2 = 10, $\overline{X}_1$ = 24.2, $\overline{X}_2$ = 23.9, s_1^2 = 10, s_2^2 = 20, δ = 0

and then $t' = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} \pm \frac{s_2^2}{n_2}}} = \dfrac{24.2-23.9}{\sqrt{\frac{10}{15} \pm \frac{20}{10}}} = \dfrac{0.3}{1.6666} = 0.18$

4. Decision: since t′ = 0.18< $t_{0.05}$ = 1.753. i.e., - 1.753 < 0.18 < 1.753 we are unable to

reject H_0 at 0.10 level of significance and that there is no strong evidence indicating

that the mean current flow is different for the two designs.

Hence the solution.

Problem #8 The following are the average weekly losses of worker hours due to accidents in 10 industrial plants before and after a certain safety program was put into operation.

Safety before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	11

Use 0.05 level of significance to test whether the safety program is effective?

[Ans. 4.03 > 1.833]

Solution: Here the parameter of interest is the mean load at failure (μ)

1. $H_0 : \mu_D = 0$ and $H_1 : \mu_D > 0$
2. $\alpha = 0.05$
3. Criterion: reject null hypothesis if $t > t_\alpha = t_{0.05} = 1.833$ is the value of t = 0.05 with 10 – 1 = 9 degrees of freedom where t = $\frac{\overline{D}-\delta}{\left(S_D \ / \sqrt{n}\right)}$ where S_D and $\overline{D}$ the s.d and mean of the differences, $\delta = \mu_1 - \mu_2$.
4. **calculations:** the differences are 9, 13, 2, 5, -2, 6, 6, 5, 2 and 6 their mean $\overline{d} = 5.2$, s.d = $S_D = 4.08$, $\delta = 0$

$$t = \frac{\overline{X} - \mu_0}{\left(s/\sqrt{n}\right)} = \frac{5.2 - 0}{\left(4.08/\sqrt{10}\right)} = \frac{5.2}{1.2903} = 4.03$$

5. Decision: since $t = 4.03 > t_{0.05} = 1.833$. Hence H_0 must be REJECTED at 0.05 level of significance and the industrial safety program is effective.

Hence the solution.

Problem #9 In random sample, 136 of 400 persons given a flu vaccine experienced some discomfort. Construct a 95% confidence interval for the true proportion of persons who will experience some discomfort from the vaccine? [Ans. 0.29 < p < 0.39]

[note : Max. Error of estimate $E = Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$]

Solution: Given n = 400, x = 136, p = x/n = 136/400 = 0.34, $\alpha = 0.05$, $z_{\alpha/2} = z_{0.05} = 1.96$

Substituting in the confidence interval formula, we get

100(1 - α) % large sample confidence interval for population parameter p

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

$$0.34 - 1.96\sqrt{\frac{0.34\ X\ 0.66}{400}} < p < 0.34 + 1.96\sqrt{\frac{0.34\ X\ 0.66}{400}}$$

$$0.294 < p < 0.386$$

By rounding to two decimals 0.29 < p < 0.39. Hence the solution.

Problem #10 In a sample survey conducted in a large city 136 of 400 persons answered yes to the question whether their city's public transportation is adequate. With 99% confidence what can be say about the maximum error if $\frac{x}{n} \equiv \frac{136}{400} \equiv 0.34$ is used as an estimate of the corresponding true proportion? [Ans. 0.061]

Solution:

Maximum error of estimate

$$\mathbf{E} = z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

with the observed value x/n substituted for p we obtain an estimate of E.

So E $= 2.575 \sqrt{\frac{p(1-p)}{n}} = \mathbf{2.575} \sqrt{\frac{0.34 X\ 0.66}{400}} \equiv 0.061$

Hence the solution.

Introduction:

The primary purpose of a statistical study is to draw conclusions about the population (parameters) based on sample (information) drawn from the population. In previous chapter, we studied how the sampling theory can be employed to obtain information about samples drawn from a known population.

The theory of statistical inference can be divided into two major areas:
(i) Estimation of parameters (ii) Test of Hypothesis

A study of either type of inferences about a population may lead to correct conjectures about the population. Procedure of estimating a population (parameter) by using sample information is referred as Estimation.

Procedures which enables one to decide whether to accept or reject hypothesis (the conjectures about the population) are called tests of hypothesis.

First, we will briefly discuss the Estimation procedures; we have two types of Estimation procedures (i) Point Estimation (ii) Interval Estimation.

Point Estimation:

Definition: A point estimate of some population parameter θ is a single numerical value.

Clearly, a question will arises about the statistic for estimating the population parameter θ.

Definition: A point estimator is a statistic for estimating a population parameter θ and will be denoted by $\hat{\theta}$.

Example: Consider the problem of point estimation of the (population) mean. Clearly, we need a single logical statistic that can be used to estimate the population mean (μ). The statistic chosen will be called a point estimator for μ because when it is evaluated for a given sample it yields a single number or a point on the real number line. The single number obtained is called a point estimate for μ.

Notice that the term estimate refers to the number obtained when the statistic is evaluated. An estimator is a random variable; an estimate is a number.

Here the logical estimator for μ is the sample mean ($\overline{X}$) i.e., the aveage of a sample drawn from that population. Hence $\hat{\mu} = \overline{X}$.

Properties of Estimators:

i)An estimator should be 'close' in some sense to the true value of unknown parameter.

ii) Formally, we say that $\hat{\theta}$ is an unbiased estimator of θ if the mean or expected value of $\hat{\theta}$ =θ.

iii) This is equivalent to saying that the mean of the probability distribution of $\hat{\theta}$ (or the mean of the sampling distribution of $\hat{\theta}$) is equal to θ.

Definition: A statistic or point estimator $\hat{\theta}$ is said to be an unbiased estimator or its value an unbiased estimate, if and only if the mean of the sampling distribution of the estimator equals θ, mean ($\hat{\theta}$) = E($\hat{\theta}$) = θ.

Definition: if the estimator is not unbiased, then the difference E($\hat{\theta}$) - θ is called the bias of the estimate $\hat{\theta}$.

When the estimator is unbiased, then E($\hat{\theta}$) - θ = 0, i.e., the bias is zero. Thus we call a statistic unbiased if on the average its value will equal to the parameter it is supposed to estimate. Also it is customary to apply the term "statistic" to both estimates and estimators. The property of unbiasedness is one of the more desirable properties in point estimation, does not necessarily indicate the close proximity of the estimator to the parameter.

As we said above, about the "closeness", closeness can be expressed in terms of (i) the bias (ii) the standard deviation (or) Variance of the estimator.

Example: The mean $\overline{X}$ and the variance $\hat{s}^2$ are unbiased estimators of the population mean μ and variance σ^2 , since E($\overline{X}$) = μ, E($\hat{s}^2$) = σ^2. The values of $\overline{X}$, $\hat{s}^2$ are called unbiased estimates. But $\hat{s}$, the sample standard deviation is a biased estimator of the population standard deviation. For large samples this bias is negligible. Sometimes there are several unbiased estimators for estimating a population parameter.

Ex.: sample mean ($\overline{X}$), sample median ($\hat{X}$) are unbiased estimators of μ,since E($\overline{X}$)= μ, E($\hat{X}$) = μ. Since there is not a unique unbiased estimator. We can not rely on the property of un-biasedness alone to select our 'best' estimator. At this situation, we need a further criterion for deciding which of several unbiased estimators is 'best' for estimating a given parameter.

Example: the sampling distribution of the means and that of medians have the same mean (μ). However the variance of the sampling distribution of means $(\frac{\sigma^2}{n})$ is smaller than the variance of the sampling distribution of the medians $\left(1.5078 X \frac{\sigma^2}{n}\right)$. Thus it is the mean ($\overline{X}$) closer to μ than the median ($\hat{X}$) is to μ.

Definition: More Efficient Unbiased Estimator [MEUE]:

A statistic $\hat{\theta}_1$ is said to be a more efficient unbiased estimator of the population parameter θ than the statistic $\hat{\theta}_2$ if (i) $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ; (ii) The variance of the sampling distribution of the first estimator $\hat{\theta}_1$ is less than (<) of the second estimator $\hat{\theta}_2$.

Example: The sample mean ($\overline{X}$) having less variance than the sample median ($\hat{X}$) is the more efficient unbiased estimator of the population mean (μ) (from the discussion in the previous example)

Inference concerning Means:

When we use a sample mean to estimate the mean of a population, along with a method of estimation having desirable properties, the chances are less, that the estimate will actually equals μ. Hence it would seem desirable to accompany such a point estimate of μ with some statement as to how close we might reasonably expect the estimate to be. The error, is the difference between the estimator and the quality it is supported to estimate. To examine this error consider, for large n, $\dfrac{\overline{X}-\mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$ a random variable having approximately the standard normal distribution.

Consequently, $P\left(-z_{\alpha/2} \le \dfrac{\overline{X}-\mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} \le z_{\alpha/2}\right) = 1 - \alpha \Rightarrow P\left(\dfrac{|\overline{X}-\mu|}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} \le z_{\alpha/2}\right) = 1 - \alpha$

$$\Rightarrow P\left(|\overline{X}-\mu| \le \frac{\sigma}{\sqrt{n}}.z_{\alpha/2}\right) = 1 - \alpha$$

Where $z_{\alpha/2}$ is such that the normal curve area to its right equals α/2. As we know that $|\overline{X}-\mu|$ is the error in estimating μ by the unbiased estimator sample mean $\overline{X}$. Now let E stands for the maximum of these values of $|\overline{X}-\mu|$ able to define maximum error of estimate with probability 1 - α.

Definition: maximum error of estimate (large sample σ known) $\boxed{E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$

If we intend to estimate μ with the mean of a large random sample, n ≥ 30 we can assert with probability 1 - α that the error $|\overline{X} - \mu|$ will be at most $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$. The most widely used values for (1 - α are 0.95 and 0.99) α are 0.05 and 0.01, and the corresponding values of $z_{\alpha/2}$ are

$z_{0.025} = 1.96$ and $z_{0.005} = 2.575$.

Definition: the formula for E can also be used to determine the sample size that is needed to attain a desired degree of accuracy. Suppose that, we wish to use the mean of a large random sample to estimate the mean of a population and we want to be able to assert with the probability (1 - α) the error would be at most some prescribed quantity E, then $\boxed{\text{Sample Size } = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2}$ to use this formula, we need 1 - α, E and σ.

Problem # an industrial engineer intends to use the mean of a random sample of size n = 150 to estimate the average mechanical aptitude (as measured by a certain test) of assembly line workers in a large industry. If, on the basis of experience, the engineer can assume that σ = 6.2 for such date, what can he assert with probability 0.99 about the maximum size of his error?[Ans. 1.30]

Solution: Substituting n = 150, σ = 6.2 and $Z_{0.005} = 2.575$ into the proceeding formula for E, we get $\boxed{E = 2.575\left(\frac{6.20}{\sqrt{150}}\right) = 1.30}$.Hence the solution.

Problem # A research worker wants to determine the average time it takes a mechanic to rotate the tires of a car, and she want to be able to assert with 95% confidence that the mean of her sample is off by at most 0.50 minute. If she can presume from past experience that σ = 1.6 minutes, how large a sample will she has to take? [Ans. 39.3~40]

Solution: substituting E = 0.50, σ = 1.6 and $Z_{0.025} = 1.96$ into the formula for n, we get $\boxed{n = \left(\frac{1.96 \, X \, 1.6}{0.50}\right)^2 = 39.3 \approx 40}$. Hence the solution.

Definition: If we intend to estimate μ with the mean of a small random sample, n < 30 we can assert with probability 1 - α that the error $|\overline{X} - \mu|$ will be at most $t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$. When $\overline{X}$ and s become available, we assert with (1 - α) 100% confidence that the errors made in using $\overline{X}$ to estimate μ is at most as defined below:

Definition: Maximum error of estimate (small sample, σ unknown) $\boxed{E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}}$

Problem # In six determinations of the melting point of tin, a chemist obtained a mean of 232.26^0 C (degree centigrade) with a standard deviation of 0.14^0 C. If he uses this mean as the actual melting point of tin, what can the chemist assert with 98% confidence about the maximum error?

Solution: substituting n = 6, s = 0.14 and $t_{0.01}$ 3.365 for n – 1 = 5 degrees of freedom into the formula for E, we get $\boxed{E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 3.365\ (0.14/\sqrt{6}) = 0.19}$.

Thus the chemist can assert with 98% confidence that his value of the melting point of tin is off by at most 0.19^0 C. Hence the Solution.

Interval Estimation:

By using point estimation, we may not get desired degree of accuracy in estimating a parameter. Therefore, it is better to replace point estimation (single numerical value) by interval estimation i.e., an estimate of a population parameter given by two numbers between which the population parameter may be supposed to lie, with a reasonable degree of certainty.

Definition: An interval estimation of an unknown parameter θ is an interval of the form L ≤ θ ≤ U, where the end points L and U depend on the numerical value of the statistic θ for a particular sample on the sampling distribution of $\hat{\theta}$.

Example: Consider the problem of estimating the average IQ of a very large group of students on the basis of a random sample, e might arrive at the interval estimate 109 ≤ μ ≤ 117 on the basis of sample mean $\overline{X}$ = 113 and sampling distribution of $\overline{X}$.

The advantage of an interval estimate over a point estimate is that the interval estimate is constructed in such a way that we can assess the confidence that the interval contains the parameter. For this reason, interval estimators are called “Confidence intervals”.

Definition 100(1 - α)% confidence interval:

A 100(1 - α)% confidence interval for a parameter θ is an interval of the form [L, U] such that P(L ≤ θ ≤ U) ≈ 1- α, 0 < α < 1 regardless of the actual value of θ.

The quantities L and U are called the lower and upper confidence limits respectively, and 1- α is called the confidence coefficient or the degree of confidence. For example or instance when α = 0.01 the confidence coefficient is 0.99 and we get 99% confidence interval.

Confidence Interval for the Mean (σ known)

Suppose that a population has unknown mean μ and known variance σ^2. A random sample of size n ($\geq$ 30) is taken from this population. The sample mean $\overline{X}$ I a reasonable point estimator of the unknown mean μ. A 100 (1 - α)% confidence interval on μ can be obtained by considering the sampling distribution of the sample mean $\overline{X}$. The sampling distribution of $\overline{X}$ is normal if the parent population is normal and approximately normal if the conditions of the central limit theorem (C.L.T) are met.

The mean of $\overline{X}$ is μ and the variance is $\frac{\sigma^2}{n}$.

Therefore the distribution of the statistic $\boxed{Z = \frac{(\overline{X}-\mu)}{\frac{\sigma}{\sqrt{n}}}}$ is a standard normal distribution. From the distribution of Z.

$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha \quad \Rightarrow \; P\left(-Z_{\alpha/2} \leq \frac{(\overline{X}-\mu)}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow \; P\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

The lower and upper limits of the inequalities in the above equation are the lower and upper confidence limits L and U respectively.

Definition: Large – sample confidence interval for μ and σ known:
If $\overline{x}$ is the sample mean of a random sample of size n from a population with known variance σ^2, a 100(1- α)% confidence interval on μ is given by

$$\boxed{\overline{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

Where, $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

The above confidence interval formula is exactly only for random samples from normal populations. But for large samples it will generally provide good approximations.

Note: since σ unknown is many practical applications therefore we have to consider the sample standard deviation s as approximation for σ.

Problem: A random sample of size n = 100 is taken from a population with σ = 5.1. Given that the sample mean is $\bar{x}$ = 21.6. Construct at 95% confidence interval for the population mean μ?

Solution: given n = 100, $\bar{x}$ = 21.6, σ = 5.1, 1 - α = 0.05 and $Z_{\alpha/2} = Z_{0.05}$ = 1.96

Substituting into the confidence interval formula $\boxed{\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$

we get $$\boxed{21.6 - 1.96\left(\frac{5.10}{\sqrt{100}}\right) \le \mu \le 21.6 + 1.96\left(\frac{5.10}{\sqrt{100}}\right)}$$

$\Rightarrow$ 20.6 < μ < 22.6. Therefore thus, we can assert with 95% confidence that the mean μ lies between the interval (20.6, 22.6).

Hence the solution.

Problem: Consider n = 80, $\bar{x}$ = 18.85 and s^2 = 30.77 (s = 5.55), construct a 99% confidence interval for the mean.

Solution: Substituting into the confidence interval formula with s = 5.55 in place of σ, we get

$$\boxed{18.85 - 2.575\left(\frac{5.55}{\sqrt{80}}\right) \le \mu \le 18.85 + 2.575\left(\frac{5.55}{\sqrt{80}}\right)}$$

$\Rightarrow$ 17.25 < μ < 20.45

We are 99% confident that the interval from 17.25 to 20.45 contains the true average μ.

Hence the Solution.

Confidence Interval for μ (σ unknown):

For small n(< 30) and when the sample is from a normal population. Thus by using t-distribution, we get

Definition: Small–sample confidence interval for μ of a Normal Distribution withσ un-known:

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, then a 100(1 - α)% confidence interval on μ is given by

$$\boxed{\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}}$$

Where, $t_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the t-distribution with n – 1 degrees f freedom. The discussion to get the above definition is similar to that of large sample confidence interval simply by replacing Z-distribution by t-distribution.

The above formula applies to samples from normal populations. It is moderately important for small samples.

Problem: The mean weight loss of n = 16 grinding balls after a certain length of time in mill slurry is 34.2 grams. With a standard deviation of 0.68 grams. Construct a 99% confidence interval for the true mean weight loss of such grinding balls under the stated conditions.

[Ans. $2.93 < \mu < 3.92$]

Solution: Substituting n = 16, $\bar{x} = 3.42$, s = 0.68 and $t_{0.005} = 2.947$ for n – 1 = 15 degrees of freedom into the small sample confidence interval formula for μ, we get

$$\boxed{\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

$$\boxed{3.42 - 2.947\left(\frac{0.68}{\sqrt{16}}\right) \leq \mu \leq 3.42 + 2.947\left(\frac{0.68}{\sqrt{16}}\right)}$$

$$\Rightarrow \quad 2.92 < \mu < 3.92$$

We are 99% confident that the interval from 2.92 to 3.92 grams contains the mean weightloss. Hence the solution.

Bayesian Estimation

In general, the parameters which are to be estimated are unknown constants. But, in recent years, the parameters in methods of inference are looked upon as random variable having prior distributions.

These methods are called Bayesian methods also, this methods of estimation called Bayesian estimations, combine sample information with prior distribution of the parameter, give posterior distribution of the parameter. Once a posterior distribution of a parameter has been obtained, it can be used to make estimates of the parameter (or it can be used to make probability statements about the parameter).

To illustrate this concept of Bayesian estimation for the mean, consider that $\bar{x}$ is the mean of a random sample of size "n" from a normal population with the known variance σ^2, and the prior distribution of μ is a normal distribution with the mean μ_0 and the variance σ_0^2 . then the posterior distribution of μ can be approximated by a normal distribution with :

Mean and standard deviation of posterior distribution
$\mu_1 \; \frac{n\bar{x}\sigma_0^2 \pm \mu_0\sigma^2}{n\sigma_0^2 \pm \sigma^2} \; and \; \sigma_1 \equiv \sqrt{\frac{\sigma^2\,\sigma_0^2}{n\sigma_0^2 \pm \sigma^2}}$

Example # Suppose we have a sample of size n 10 from a normal distribution with unknown mean μ and variance σ^2 4. Assume that the prior distribution for μ is normal with mean $\mu_0 = 0$ and variance $\sigma_0^2 = 1$. If the sample mean is 0.75, then the mean of posterior distribution estimated by Bayesian formula

$$\mu_1 \equiv \frac{10(0.75)(1) \pm 0(4)}{10(1) \pm 4} \equiv \frac{7.5}{14} \cong 0.5357$$

and the standard deviation

$$\sigma_1 = \sqrt{\frac{4(1)}{10(1) \pm 4}} \equiv \frac{2}{\sqrt{14}} \equiv \frac{2}{3.7416} \equiv 0.5345.$$

Thus the mean and standard deviation of posterior distribution are 0.5357 and 0.5345 respectively.

Bayesian interval of μ : A(1 - α) 100% Bayesian interval for μ is :

$$\mu_1 - Z\alpha/2.\frac{1}{\sigma_1} < \mu < \mu_1 + Z\alpha/2.\frac{1}{\sigma_1}$$

The main object of the sampling theory is the study of the Tests of hypothesis/Tests of significance.

Definition: There are many problems, in which, rather than estimating the value of a parameter (in engineering, science and management) we need to decide whether to accept or reject a statement about the parameter. This statement is called Hypothesis and the decision making procedure about the hypothesis is called hypothesis testing.

Example: (i) A drug chemist is to decide whether a new drug is really effective in curing a disease.
(ii) A quality control manager is to determine whether a process is working properly
(iii) A statisitician has to decide whether a given coin is biased
Such decisions are called statistical decisions or simply decisions.

Definition: To arrive at decisions about the population on the basis of sample information, we make assumptions or guesses about the population parameters involved. Such an assumption or statement is called a statistical hypothesis which may or may not be true.
The procedure which enables us to decide on the basis of sample results whether a hypothesis is true or not, is called Test of hypothesis or test of significance.

Example: (i) The majority of men in the city are smokers.
(ii) The teaching methods in both the schools are effective.

Steps involves in Test of Hypothesis:

Step 1: Statement or assumption of hypothesis –

Type I : Null hypothesis – for applying tests of significance, we first set up a hypothesis a definite statement about the population parameter. Such a hypothesis is usually a hypothesis of no-difference, is called Null Hypothesis.

Definition: A null hypothesis is the hypothesis which asserts that there is no significant difference between the statistic and the population parameter and whatever observed difference is there, is merely due to fluctuations in sampling fromt he same population. It is always denoted by H_0.

Example: A single statistic, H_0 will be that the sample statistic does not differ significantly from the hypothetical parameter value and in the case of two statistics (H_0) will be that the sample statistics do not differ significantly.

Type –II Alternate Hypothesis: Any hypothesis which contradicts the Null hypothesis is called an Alternate Hypothesis, usually denoted by H_1. The two hypothesis H_0 and $H_1$are such that if one is true, the other is false and vice versa.

Example: if we want to test the null hypothesis that the population has a specified mean μ_0 say i.e., $H_0 : \mu = \mu_0$ then the alternate hypothesis would be

(i) $H_1 : \mu \neq \mu_0 \Rightarrow$ either $\mu > \mu_0$ or $\mu < \mu_0$ or

(ii) $H_1 : \mu > \mu_0$ or

(iii) $H_1 : \mu < \mu_0$

The alternate hypothesis is very important to decide whether we have to use a single – tailed (right or left) or two-tailed test.

Step 2: Specification of the level of significance –

Level of significance is the size of test. Level of significance denoted by a α is the confidence with which we rejects or accepts the null hypothesis H_0 i.e., it is the maximum possible probability with which we are willing to risk an error in rejecting H_0 when it is true.

Step 3: Identification of the test statisitic

There are several tests of significance namely z, t , F etc. First we have to select the right test depending on the nature of the information given in the problem. Then we construct the test criterion and select the appropriate probability distribution.

Step 4: Making decision

We compute the value of the appropriate statistic and ascertain whether the computed value falls in acceptance or rejection region depending on the specified level of significance. In finding the acceptance or rejection region we have to use critical value decision is taken for accepting or rejecting H_0. If the computed value < critical value, we accept H_0, otherwise we reject H_0.

Error sampling: The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of the sample results. In practice we decide to accept or reject the lot after examining a sample from it. As such we have two types of errors.

(i) Type I error: reject H_0 when it is true. If the null hypothesis H_0 is true but it rejected by test procedure, then the error made is called type I error or α error.

(ii) Type II Error: Accept H_0 when it is wrong i.e., accept H_0 when H_1 is true. If the null hypothesis is false but it is accepted by test, then error committed is called type II error or β error.

if we write P(Reject H_0 when it is true) = P (type I error) $=\alpha$

and P(Reject H_0 when it is false) = P (type II error) $=\beta$

then α and β are called sizes of type I and type II errors respectively

i.e., α = P(rejecting a good lot)

β = P(accepting a bad lot)

The sizes of type I and type II errors also known as producer's risk and consumer's risk respectively.

1. **Critical Region:** A region corresponding to a statistic ‘t’ in the sample space S which leads to the rejection of H_0 is called critical region and rejection region. Those region which lead to the acceptance of H_0 give us a region called acceptance region.

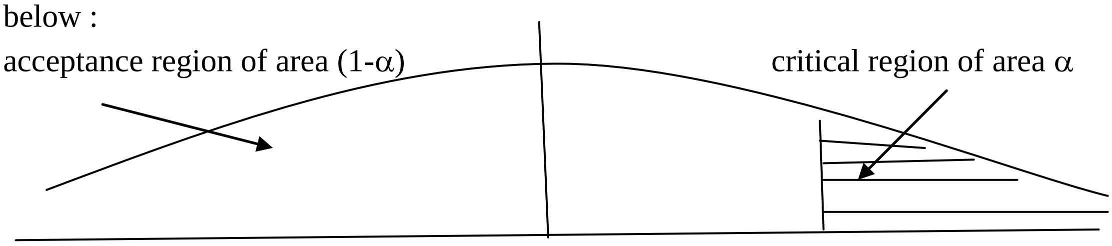
2. **Critical values or Significant values:** it is the value of the test statistic θ, which divides the area under the probability curve into critical region, for given level of significance α. It is also known as significance value.

3. **Two tailed test at level of significance ‘α’**

Right one tailed test:

When H_1 is of the greater than type : $H_1 : \mu > \mu_0$ or $\sigma_1^2 > \sigma_2^2$, then the complete critical region of area α lies on the right side tail of the probability density curve as shown below :

acceptance region of area (1-α) critical region of area α



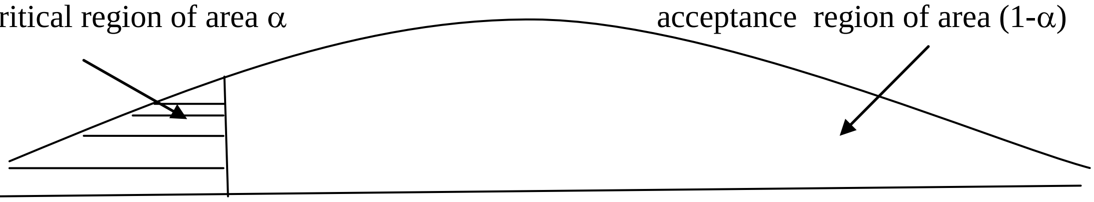
θ_α

One sided right tail test

Left one tailed test:

When H_1 is of the less than type-that $H_1 : \mu < \mu_0$ or $\sigma_1^2 < \sigma_2^2$, then the complete critical region of area α lies on the left side tail of the probability density curve as shown below :

critical region of area α acceptance region of area (1-α)

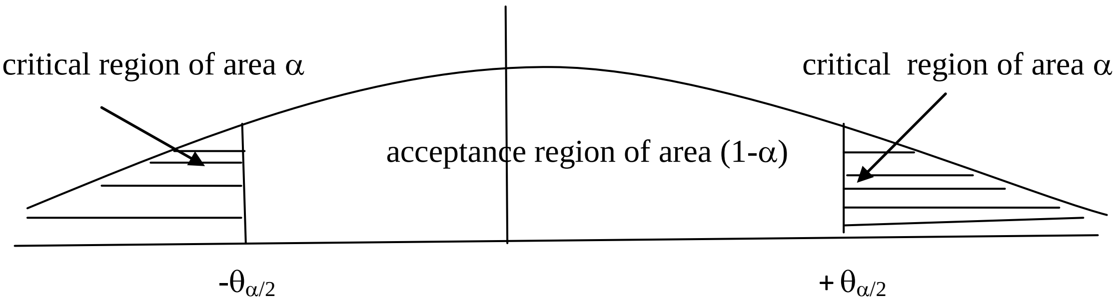


θ_α

One sided left tail test

Two tailed tests:

When H_1 is not equals type that $H_1 : \mu_1 \neq \mu_2$ or $\sigma_1 \neq \sigma_2$, or $\mu_1 \neq \mu_0$ then the critical region of area α lies on both sides of the right and left tails of the curve such that the



Two sided right and left of tails test

critical region $\frac{\alpha}{2}$ lies on the right tail and critical region of area $\frac{\alpha}{2}$ lies on the left tail, as shown above figure :

Guidelines for testing hypothesis systematically as outlined in the following five steps:

Step 1. We formulate a null hypothesis and an appropriate alternate hypothesis which we accept when the null hypothesis must be rejected.

Step 2. We specify the probability of a type I error; if possible, desired or necessary, we may also specify the probabilities of type II errors for particular alternatives after the H_0 , H_1 and α have been specified, the remaining steps are:

Step 3. Based on the sampling distribution of an appropriate statistic, we construct a criterion for testing the null hypothesis against the given alternative.

Step 4. We calculate from the data the value of the statistic on which the decision is to be based.

Step 5. We decide whether to reject the null hypothesis, whether to accept it, or whether to reserve judgement.

Note: Null hypothesis should be framed as a simple hypothesis, since α is dependent on the null hypothesis.

General procedure for testing hypothesis in eight steps:

1. From the problem statement content, identify the parameter of interest
2. State the null hypothesis, H_0
3. Specify an appropriate alternate hypothesis H_1
4. Choose a significance level α
5. State an appropriate test statistic
6. State the rejection region for the statistic
7. Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute the value
8. Decide whether or not H_0 should be rejected and report that in the problem context.

Hypothesis testing procedures for many practical problems we will consider here one by one.

1. **Hypothesis concerning one mean(with known variance σ^2 : large sample)**
2. **Hypothesis concerning one mean(with unknown variance σ^2 : small sample)**
3. **Inferences concerning two means (σ_1 and σ_2 known or Large sample)**
4. **Hypothesis concerning two means (σ_1 and σ_2 unknown or small sample)**
5. **Small sample confidence interval concerning difference between two means**

Hypothesis concerning one mean(with known variance σ^2 : large sample)

Statistic for test concerning mean (large sample) Z = $\dfrac{\overline{X}-\mu_0}{\left(\sigma/\sqrt{n}\right)}$

Critical regions values of Z for which reject the null hypothesis $\mu=\mu_0$

Alternate hypothesis	Rejection null hypothesis if;	Tests level of significance→	α = 0.01	α = 0.05
$\mu < \mu_0$	$Z < -Z_\alpha$	Left sided	Z < - 2.33	Z < - 1.645
$\mu > \mu_0$	$Z > Z_\alpha$	Right sided	Z > 2.33	Z > 1.645
$\mu \neq \mu_0$	$Z < -Z_\alpha$ Or $Z > Z_{\alpha/2}$	Two sided alternative	Z < - 2.575 Or Z > 2.575	Z < - 1.96 Or Z > 1.96

Definition: P-value

The P-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0.

If the customary to call the test statistic and the data significant when the null hypothesis H_0 is rejected. So, P-value as the smallest level α at which the data are significant. Once P-value is known, we can determine how significant the data are. For the forthcoming normal distribution tests, it is relatively easy to compute the P-value.

If Z is the compound value of the test statistic, then the P-value is given as below:

$$P = \begin{cases} 2(1 - F(|Z|) & for\ H_0 : \mu \equiv \mu_0, H_1 : \mu \neq \mu_0 \\ 1 - F(Z) & for\ H_0 : \mu \equiv \mu_0, H_1 : \mu > \mu_0 \\ F(Z) & for\ H_0 : \mu \equiv \mu_0, H_1 : \mu < \mu_0 \end{cases}$$

Where F(Z) is normal cumulative distributive function, recall F(Z) = P(Z $\leq$ z), where Z ~ N(0,1).

Problem #1 Suppose that we want to test on the basis of n = 35 determinants and at the 0.05 level of significance whether the thermal conductivity of a certain kind of cement brick is 0.340, as has been claimed. From information gathered in similar studies, we can expect that the variability of such determinations is given by σ = 0.010 and suppose that the mean of 35 determinations is 0.343?

Solution: Here the parameter of interest is the mean of thermal conductivity (μ)

6. $H_0 : \mu = 0.340$ and $H_1 : \mu \neq 0.340$
7. α= 0.05
8. Criterion: reject null hypothesis if Z < - 1.96 or Z > 1.96 where Z = $\boxed{\dfrac{\overline{X} - \mu_0}{\left(\sigma/\sqrt{n}\right)}}$
9. **calculations:** since n = 35, σ=0.010, $\overline{X}$ = 0.343, $\mu_0 \equiv 0.340$

$$Z = \frac{\overline{X} - \mu_0}{\left(\sigma/\sqrt{n}\right)} = \frac{0.343 - 0.340}{\left(0.010/\sqrt{35}\right)} \equiv 1.77$$

10. decision: since Z = 1.77 < 1.96 hence H_0 CAN'T BE REJECTED.

To find P-value: Z = 1.77 and H_1 is two tailed so P-value is

$$P = 2(1 - F(1.77)) = 2(1 - 0.9616) = 0.07868$$

Therefore P $\geq \alpha$ = 0.05 H_0 can't e rejected, agreeing our earlier result.

Hence the solution.

Problem #2 A trucking suspects the claim that the average life time of certain tires is atleast 28,000 miles. To check the claim, the firm puts 40 of these tires on its trucks and gets a mean life time of 27,463 miles with a standard deviation of 1,348 miles. What can it conclude if the probability of a type I error is to be at most 0.01?

Solution: Here the parameter of interest is the life time of certain tires (μ)

5. $H_0 : \mu > 28000$ and $H_1 : \mu < 28000$
6. $\alpha \leq 0.01$
7. Criterion: reject null hypothesis if Z < - 2.33 or Z > 2.33 where Z = $\frac{\overline{X} - \mu_0}{\left(\sigma/\sqrt{n}\right)}$ with σ replaced by s since it is of type I error, level of significance is 0.01
8. **calculations:** since n = 40, s=1348, $\overline{X}$ = 27643, $\mu_0 \equiv 28000$

$$Z = \frac{\overline{X} - \mu_0}{\left(s/\sqrt{n}\right)} = \frac{27643 - 28000}{\left(1348/\sqrt{40}\right)} \equiv -2.52$$

9. decision: since Z = -2.52 < - 2.33 hence H_0 must be REJECTED at 0.01 level of significance.

Hence the solution.

Hypothesis concerning one mean (σ^2 unknown: small sample):

If the sample size is small and σ is unknown , the test just described can not be used. Then, we must make an assumption about the form of the underlying distribution in order to obtain a test procedure. A reasonable assumption in many cases is that the underlying distribution is normal. Therefore, suppose that the population of interest has a normal distribution with unknown mean μ and variance σ^2, assume that a random sample of size n, and let $\overline{X}$ and s^2 be sample mean and variance respectively.

Now we wish to test the hypothesis $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$ or $\mu > \mu_0$ or $\mu < \mu_0$

Then we can make use of the theory discussed in earlier sections, and base the test of $H_0 : \mu = \mu_0$ on the following statistic:

Statistic for small – sample test concerning mean

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

which is a random variable having t-distribution with n-1 degrees of freedom. The criteria or critical values for t-test based on this statistic are like those in the table of Z-statistic with z, z_α and $z_{\alpha/2}$ are replaced by t, t_α and $t_{\alpha/2}$ respectively.

Critical regions for testing $H_0 : \mu = \mu_0$ (small – sample, σ^2 unknown)

Alternate Hypothesis	Reject H_0 if
$\mu < \mu_0$	$t < - t_\alpha$
$\mu > \mu_0$	$t > t_\alpha$
$\mu \neq \mu_0$	$t < - t_{\alpha/2}$ or $t > t_{\alpha/2}$

Inferences concerning two means (σ_1 and σ_2 known or Large sample):

In applied research, there are many problems in which we are interested in hypothesis concerning differences between the means of two populations. Now, we devote this discussion to tests concerning the difference between two means. Formulating the problem more generally, let us suppose that we are dealing with independent random sample of size n_1 and n_2 from two normal populations having means μ_1 and μ_2 and the known variances σ_1^2 and σ_2^2 And that we want to test the null hypothesis $\mu_1 - \mu_2 = \delta$, where δ is a given constant, against one of the alternatives $\mu_1 - \mu_2 \neq \delta$, $\mu_1 - \mu_2 > \delta$ or $\mu_1 - \mu_2 < \delta$. The test procedure is based on the distribution of the difference in the sample mean $\overline{X_1} - \overline{X_2}$ with $\mu_1 - \mu_2$ and the variance $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. That is $\overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ and it can be based on the statistic:

Statistic for test concerning difference between two means

$$\mathbf{Z} = \frac{\left(\overline{X}_1 - \overline{X}_1\right) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Which is a random variable having the standard normal distribution. However, the above statistic can also be used, when we deal with large independent random sample from populations with unknown variances which may not even be normal, by invoking central limit theorem with s_1 substituted for σ_1 and s_2 substituted for σ_2.

The critical regions for testing the null hypothesis $\mu_1 - \mu_2 = \delta$ normal populations with σ_1 and σ_2 known or large samples

Alternate Hypothesis	Reject H_0 if
$\mu < \mu_0$	$Z < -z_\alpha$
$\mu > \mu_0$	$Z > z_\alpha$
$\mu \neq \mu_0$	$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$

Although δ can be any constant, but in most of the problem its value is zero and we test the null hypothesis of ‘no difference’, namely $H_0 : \mu_1 = \mu_2$.

Probability and statistical Applications

Estimation theory Unit III by Dr. N V Nagendram

Q1. Define types of Estimation.

Q2. Properties of good estimator

Q3. Define point estimation, interval estimation and Bayesian estimation.

Q4. Using the mean of a random sample of size 150 to estimate the mean mechanical aptitude of mechanics of a large workshop and assuming $\sigma = 6.2$. What can we assert with 0.99 probability about the maximum size of the error?

Q5. Assuming that $\sigma = 20.0$ how large a random sample between to assert with probability 0.95, that the sample mean will not differ from the true mean by more than 3.0 points?

Q6. The pulse rate of 50 yoga practitioners decreased on an average by 20.2 beats per minute with s.d of 3.5.

(i) If $\bar{x} = 20.2$ is used as a point estimate of the true average what can we assert with 95% confidence about max. error?

(ii) Construct 99% and 95% confidence intervals for true average?

Q7. Find 95% confidence limits (or intervals) for the mean of a normally distributed population from which the following sample was taken 15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

Q8. Calculate μ_1, σ_1 for the posterior distribution then find the Bayesian interval at 95% if the random sample size is 80 and $\bar{x} = 18.85$, S = 5.55 using S for s.d of population σ?

Q9. A random of sample of 100 teachers has a many weekly salary of Rs. 487/- with a s.d of Rs. 48/- with what degree of confidence can we assert the average weekly salary of all teachers is between 472 and 502?

**** **** **** **** ****

Probability and statistical Applications

Estimation theory **Unit III** **by Dr. N V Nagendram**

01. If the size of the sample is 200 and of population is 2000, the correction factor
   a) 0.8 b) 0.7 c) 0.9 d)0.99 Ans. C

02. If the size of the sample is 5 and of population is 2000, the correction factor
   a) 0.712 b) 0.888 c) 0.999 d)0.566 Ans. C

03. If the size of the sample is 121 and s.d of the population is 2, the standard error of sample mean is
   a) 0.182 b) 0.171 c) 0.23 d)0.32 Ans. A

04. If a sample is taken from an infinite population and sample size is increased from 25 to 100 the effect on standard error is
   a) Same value b) c) divided by 2 d)multiplied by 4 Ans. C

05. If a sample is taken from an infinite population and if the sample size is decreased from 800 to 200 the effect on standard error is
   a) multiplied by 2 b) divided by 2 c) multiplied by 3 d)divided by 3 Ans. C

06. If a sample size 64 is taken from a population whose standard deviation is 4 the probable error is
   a) 0.125 b) 0.337 c) 0.432 d)0.532 Ans. C

07. A random sample of size 64 is taken from an infinite population having mean 50 and variance 25. The probability that $\bar{x}$ is less than 51.5i.e, ($\bar{x} <$ 51.5)
   a) 0.9918 b) 0.821 c) 0.7 d)0.521 Ans. A

08. $\mu = 30.5$; $r = 100$; $\bar{x} = 28.8$; $\sigma = 8.35$ then $|Z| =$
   a) 1.68 b) 2.68 c) 3.68 d)4.68 Ans. B

09. The sample observations of size 4 are 25, 28, 26, 25 the the variance of sample is
   a) 1 b) 2 c) 3 d)4 Ans. B

Probability and statistical Applications

Estimation theory   Unit III   Quiz 1(3)   by Dr. N V Nagendram

10. If the population is 2, 4, 6, 8, 10 if samples of size 2 are taken with replacement. Then the mean of the means of sampling distribution is
    a) 1   b) 2   c) 5   d) 6   Ans. D
11. A sample of size 100 is taken whose s. d is 5 what is the max. error at 95% Level of Confidence (L.C.)
    a) 0.8   b) 0.7   c) 1   d) 0.98   Ans. D
12. If the maximum error with probability 0.95 is 1.2, s. d of population is 10. Then the sample size is
    a) 267   b) 129   c) 225   d) 169   Ans. A
13. If the maximum error with probability 95% C.I is 0.1, size of the sample is 144 then s. d of population σ is
    a) 0.54   b) 0.61   c) 0.45   d) 0.72   Ans. B
14. If n = 144, σ = 4 and $\bar{x}$ = 150 then 95% confidence interval (C.I.) for μ is
    a) (139.72, 140.25)   b)(149.35, 150.65)
    c) (172.1, 182.12)   d) (179.1, 182.25)   Ans. B
15. If n = 9, s = 0.15 the max error with 0.99 probability is
    a) 0.168   b) 0.272   c) 0.324   d) 0.468   Ans. A
16. the unbiased estimator for population mean μ is
    a) A.M($\bar{x}$)   b) $\frac{1}{n}\sum x_i^2$ c) mode   d) H.M.   Ans. A
17. the unbiased estimator for population variance σ^2 is
    a) $S^2=\frac{1}{n}\sum(x_i-\bar{x})^2$ b) $S^2=\frac{1}{n-1}\sum(x_i-\bar{x})^2$
    b) c) $S=\sqrt{\frac{1}{n}\sum(x_i^2)}$ d) None   Ans. B
18. If n = 169 the variance of population is 25 and $\bar{x}$ = 50 then 99% confidence interval for μ is
    a) (42,44)   b) (49,51)   c) (36, 38)   d) (24, 26)   Ans. B
19. If n = 100, p = $\frac{51}{100}$, α = 0.01 then the 99% confidence interval for p is
    a) (0.42, 0.68) b) (0.40, 0.70)   c) (0.41, 0.69) d) None   Ans. A
20. Bayesian interval for μ is
    a) $\mu_0 \pm \sigma_0$ b) $\mu_1 \pm \sigma_1$
    c) $\mu_1 \pm Z_{\alpha/2}.\sigma_1$ d) $\mu_0\mu_1 \pm Z_{\alpha/2}\,\sigma_0.\sigma_1$ Ans. C

*****   *****   *****